Topics in Geodesic Convexity

Arpon Basu

Last updated April 8, 2024

Contents

Introduction

The importance of convex optimization can't be overstated, and not unsurprisingly, the field of convex optimization has also been very intensively researched over the decades. However, powerful as convex optimization might be, many problems of practical interest arising in optimization and machine learning are non-convex, and thus the conventional tools of convex optimization fail.

However, as it turns out, a special class of non-convex functions becomes convex once one imposes a different manifold structure on the underlying domain. Indeed, consider the function $f : \mathbb{R}^n_{>0} \mapsto \mathbb{R}$, where:

$$f(x_1, \ldots, x_n) := \ln(x_1^2 x_2^2 \cdots x_n^2 + x_n^{2n}) - \sum_{i=1}^{n} \ln(x_i)$$

f is clearly not convex. However, under the manifold structure induced by the Hessian of the function $-\sum_{i=1}^{n} \ln(x_i)$ (see Section 1.1.2), the function is *geodesically convex* (see Lemma 2.11, Lemma 2.12), and consequently, we may use tools from convex optimization to optimize this function (which arises in computing maximum entropy distributions: See [Gur06, SV17]).

Since the above function, many more non-convex functions arising in very natural contexts were found to be geodesically convex when the 'correct' Riemannian metric tensor was imposed on the manifold (instead of the usual Euclidean metric tensor). We shall see some examples in Section 3.

Thus, in this report, we shall invest ourselves in the study of convexity over Riemannian manifolds. In this process, we have referred to [Vis18, Bou23]: We would like to thank Vishnoi and Boumal for making a somewhat intimidating topic much more accessible.

The rough outline of the text is as follows: In the first chapter, we introduce the concept of geodesics, and explicitly compute them for a few manifolds of our interest. In the second chapter, we introduce and develop the theory of geodesic convex optimization, ending the chapter with an analog of gradient descent on Riemannian manifolds (see Theorem 2.19). We end the report by discussing two non-convex functions of practical interest that become geodesically convex under a suitable Riemannian metric.

The reader may also refer to the appendix for various useful facts about Riemannian manifolds that have been used liberally throughout the text.

Finally, we would also like to thank Prof. Debasish Chatterjee for his excellent lectures and notes on the rudiments of differential geometry.

§1. Pseudo-Riemannian Manifolds and Geodesics

We first define the useful notion of *frames* for manifolds in general:

Definition 1.1 (Frames)**.** Let M be a smooth n-dimensional manifold (not necessarily equipped with a metric tensor). A frame bundle on M is an ordered tuple $(\partial_1, \dots, \partial_n)$ of smooth vector fields (i.e. maps from M to TM) such that for any $p \in M$, $\{\partial_i(p)\}_{i\in[n]}$ forms a basis for T_pM.

From now on, all vectors in T_pM will be expressed in some fixed frame bundle basis. We shall not explicitly state the frame bundle basis.
We can now define a (pseudo)-Riemannian manifold.

Definition 1.2. A (pseudo)-Riemannian manifold is a smooth n-dimensional manifold M equipped with a smooth function $g : M \mapsto \mathcal{S}_n$, where $\mathcal{S}_n$ is the manifold of $n \times n$ symmetric invertible matrices.

The map g is sometimes also called the *metric tensor*: Indeed, for any $p \in M$, we have an inner product on T_pM, given by $\langle u, v\rangle_g := u^\top G(p)v$, where T_pM is identified with $\mathbb{R}^n$. Clearly, once we have an inner product, we can also define a norm, and hence a metric. However, the reader is asked to note that the metric will be non-negative only if the image of g lies in the space of positive-definite matrices (in which case we call M a Riemannian manifold).
The definition of a metric tensor also allows us to define the Christoffel symbols:

Definition 1.3 (Christoffel Symbols)**.** Fix a point p. $G(p)$ is a $n\times n$ matrix, and let the $(i, j)^{\text{th}}$ entry of $G(p)$ be denoted as g_{ij} (Note that the matrix representation of G is assumed to be in some fixed frame bundle basis). Also, denote the $(i, j)^{\text{th}}$ entry of $G(p)^{-1}$ as g^{ij}. Then we define:

$$\Gamma^k_{ij} := \frac{1}{2}\sum_{\ell=1}^{n} g^{\ell k} \cdot \left(\frac{\partial g_{\ell i}}{\partial y_j}(y) + \frac{\partial g_{\ell j}}{\partial y_i}(y) - \frac{\partial g_{ij}}{\partial y_\ell}(y)\right)$$

Once we have a pseudo-Riemannian manifold, we can define a *geodesic*. Indeed, let $\gamma : [0, 1] \mapsto M$ be a smooth map, where $[0, 1]$ should be thought of as time ('t'). We define the length of γ as:

$$L(\gamma) := \int_0^1 \sqrt{\langle\dot\gamma(t), \dot\gamma(t)\rangle_{g(\gamma(t))}}dt \tag{1.1}$$

We also define the *energy* of γ as:

$$\mathcal{E}(\gamma, \dot\gamma, t) := \frac{1}{2}\langle\dot\gamma(t), \dot\gamma(t)\rangle_{g(\gamma(t))} \tag{1.2}$$

Indeed, in Euclidean space, the (kinetic) energy of a (unit mass) particle is $v^\top v/2$, where v is the velocity of the particle. Since we are working in Euclidean space, energy was obtained from velocity by $v^\top v$. Similarly, on a general (pseudo-)Riemannian manifold, the energy of a particle with velocity $v = \dot\gamma(t)$ at a point $x = \gamma(t)$ will be given by $1/2\langle\dot\gamma(t), \dot\gamma(t)\rangle_{g(\gamma(t))}$.

Definition 1.4 (Geodesics)**.** Define the energy functional to be:

$$S(\gamma) := \int_0^1 \mathcal{E}(\gamma, \dot\gamma, t)dt = \frac{1}{2}\int_0^1 \langle\dot\gamma(t), \dot\gamma(t)\rangle_{g(\gamma(t))}dt$$

Effectively, $S(\gamma)$ denotes the work done to move the particle along the curve.
A curve $\gamma_* : [0,1] \mapsto M$, with $\gamma_*(0) = p, \gamma_*(1) = q$, is called a *geodesic* if it is a critical point of the energy functional.

Remark. A geodesic may not always exist: For example, consider the manifold $\mathbb{R}^2 - \{(0,0)\}$ equipped with the usual metric tensor. There is no geodesic between $(1,1)$ and $(-1,-1)$ on this manifold. Thus, from this point onwards, we shall only deal with pseudo-Riemannian manifolds for which there is a geodesic between any two given points.

Using the Euler-Lagrange equations to characterize geodesics yields that γ_* must satisfy the following differential equation:

$$\frac{d}{dt}\frac{\partial \mathcal{E}}{\partial \dot{\gamma}} = \frac{\partial \mathcal{E}}{\partial \gamma} \tag{1.3}$$

Simplifying the above, we obtain that a geodesic γ_* must satisfy the following series of second-order differential equations: Indeed, fix a point $p \in M$, and consider a chart (U, φ) such that $p \in U$. Define $x_* : [0,1] \mapsto \mathbb{R}^n$ as $x_* := \varphi \circ \gamma_*$. Then, for every $k \in [n]$, we have:

$$(\ddot{x}_*)_k = -\sum_{i,j=1}^{n} \Gamma^k_{ij}(x_*)(\dot{x}_*)_i(\dot{x}_*)_j \tag{1.4}$$

With this very useful formula in place, we can compute the geodesics for various manifolds of interest.

1.1. Calculating Geodesics for various Manifolds

We consider various examples of pseudo-Riemannian manifolds and investigate how geodesics look on them.

1.1.1. Euclidean Space

Consider the manifold $\mathbb{R}^n$, equipped with the metric tensor which coincides with the ordinary dot product at every point, i.e. $g_p(u,v) = u^\top v$ for all $p \in \mathbb{R}^n$. Since the metric tensor doesn't change with space, all the Christoffel symbols are 0, and we obtain $\ddot{x}_* = 0$. But the only smooth function with zero second-derivative everywhere is the line, and thus we recover the usual fact that lines are the geodesics on $\mathbb{R}^n$.

1.1.2. Positive Orthant

Consider the manifold $\mathbb{R}^n_{>0}$, whose smoothness is inherited from $\mathbb{R}^n$, i.e. we consider $\mathbb{R}^n_{>0}$ to be embedded in $\mathbb{R}^n$. However, we impose a different metric tensor now. Consider the *log-barrier* function $f : \mathbb{R}^n_{>0} \mapsto \mathbb{R}$, given by $f(x_1, \ldots, x_n) := -\sum_{i=1}^n \ln(x_i)$. Then the Hessian of f is a diagonal matrix whose diagonal entries are $x_1^{-2}, \ldots, x_n^{-2}$. Thus, for any point $p = (x_1, \ldots, x_n) \in \mathbb{R}^n_{>0}$, we define $G(p) := \mathrm{diag}(x_1^{-2}, \ldots, x_n^{-2})$. It is not too difficult to see that the Christoffel symbols are given by

$$\Gamma^k_{ij} = \begin{cases} -1/x_k & \text{if } i = j = k \\ 0 & \text{otherwise} \end{cases}$$

Eq. (1.4) then simplifies to $(\ddot{x})_k = (\dot{x})_k^2 / x_k$ for every $k \in [n]$, solving which yields $x(t) = c_1 \cdot c_2^t$ for some constants $c_1, c_2 \in \mathbb{R}^n_{>0}$, where c_2^t is defined componentwise. Now, assuming $x(0) = p, x(1) = q$, we obtain the geodesic on the positive orthant between p and q is parametrized as:

$$x(t) = (p_1(q_1/p_1)^t, \ldots, p_n(q_n/p_n)^t)$$

Remark: Pseudo-Riemannian manifolds whose metric tensor is the Hessian of some function are also known as Hessian Manifolds.

1.1.3. The Positive Definite Cone

Let $\mathbb{S}^k_{++}$ be the space of all positive-**definite** symmetric $k \times k$ matrices, i.e. if $S \in \mathbb{S}^k_{++}$, then all eigenvalues of S are strictly positive. We consider $\mathbb{S}^k_{++}$ as a submanifold of $\mathbb{R}^{k(k+1)/2}$. Note that under this embedding, the standard inner product between matrices A, B is given by $\langle A, B\rangle = \operatorname{tr}(A^\intercal B) = \operatorname{tr}(AB)$. Indeed, note that $(A^\intercal B)_{ij} = \sum_\ell a_{\ell i} b_{\ell j}$, and thus $\operatorname{tr}(A^\intercal B) = \sum_i \sum_\ell a_{\ell i} b_{\ell i}$, which is exactly what one would get if one flattens out A, B into vectors in $\mathbb{R}^{k(k+1)/2}$ and then computes the usual inner product.

If we want to "weight" this inner product by some invertible matrix X, the "weighted" inner product is given by $\langle A, B\rangle_X := \operatorname{tr}(X^{-1}AX^{-1}B)$. Indeed, recall that "re-weightings" of the usual inner product on $\mathbb{R}^n$ are given by $\langle u, Sv\rangle$, where S is some positive-definite matrix. Since S is positive-definite, we have a matrix T (not necessarily symmetric) such that $S = T^\intercal T$, and thus $\langle u, Sv\rangle = \langle Tu, Tv\rangle$. To make certain algebraic manipulations easier, we re-weigh A, B by the inverse of the matrix X, rather than the matrix X itself.

Finally, the metric tensor on $\mathbb{S}^k_{++}$ at some point $S \in \mathbb{S}^k_{++}$ is given by $g_S(U, V) := \operatorname{tr}(S^{-1}US^{-1}V)$, i.e. the matrix itself acts as the "re-weighting" agent at a given point.

We now want to calculate the geodesics of this manifold. Now, note that if g is the metric tensor for this manifold, then for any S, $g(S)$ is a $k(k+1)/2 \times k(k+1)/2$ matrix. Consequently, writing down the Christoffel symbols for this manifold becomes notationally cumbersome.

Consequently, we shall instead solve the Euler-Lagrange equations directly for this manifold. Before we begin with that, define:

$$E_{ij} = \begin{cases} e_i e_j^\intercal + e_j e_i^\intercal & \text{if } i \neq j \\ e_i e_i^\intercal & \text{otherwise} \end{cases}$$

Note that $\{E_{ij}\}_{i,j\in[k]}$ form a basis for $\mathcal{S}_k$, where $\mathcal{S}_k$ is the set of *all* symmetric matrices. Thus we can write $\gamma : [0,1] \mapsto \mathbb{S}^k_{++} \hookrightarrow \mathcal{S}_k$ as

$$\gamma(t) = \sum_{i,j\in[k]} \gamma_{ij}(t) E_{ij}$$

Consequently,

$$\frac{\partial \gamma}{\partial \gamma_{ij}} = E_{ij}$$

Now, from Eq. (1.2) (we ignore the factor of $1/2$),

$$\mathcal{E}(\gamma, \dot\gamma, t) = \operatorname{tr}(\gamma(t)^{-1}\dot\gamma(t)\gamma(t)^{-1}\dot\gamma(t))$$

Thus (by Eq. (A.1)),

$$\frac{\partial \mathcal{E}}{\partial \gamma_{ij}} = \operatorname{tr}\left(\frac{\partial \gamma^{-1}}{\partial \gamma_{ij}}\dot\gamma\gamma^{-1}\dot\gamma + \gamma^{-1}\frac{\partial \dot\gamma}{\partial \gamma_{ij}}\gamma^{-1}\dot\gamma + \gamma^{-1}\dot\gamma\frac{\partial \gamma^{-1}}{\partial \gamma_{ij}}\dot\gamma + \gamma^{-1}\dot\gamma\gamma^{-1}\frac{\partial \dot\gamma}{\partial \gamma_{ij}}\right)$$

Also, by Eq. (A.2), we have

$$\frac{\partial \gamma^{-1}}{\partial \gamma_{ij}} = -\gamma^{-1}E_{ij}\gamma^{-1}$$

Also, note that

$$\frac{\partial \dot\gamma}{\partial \gamma_{ij}} = 0$$

Thus,

$$\frac{\partial \mathcal{E}}{\partial \gamma_{ij}} = -2\operatorname{tr}\left(\gamma^{-1}E_{ij}\gamma^{-1}\dot\gamma\gamma^{-1}\dot\gamma\right) = -2\operatorname{tr}\left(E_{ij}\gamma^{-1}\dot\gamma\gamma^{-1}\dot\gamma\gamma^{-1}\right)$$

Similarly,

$$\frac{\partial \mathcal{E}}{\partial \dot\gamma_{ij}} = \operatorname{tr}\left(\frac{\partial \gamma^{-1}}{\partial \dot\gamma_{ij}}\dot\gamma\gamma^{-1}\dot\gamma + \gamma^{-1}\frac{\partial \dot\gamma}{\partial \dot\gamma_{ij}}\gamma^{-1}\dot\gamma + \gamma^{-1}\dot\gamma\frac{\partial \gamma^{-1}}{\partial \dot\gamma_{ij}}\dot\gamma + \gamma^{-1}\dot\gamma\gamma^{-1}\frac{\partial \dot\gamma}{\partial \dot\gamma_{ij}}\right) = \operatorname{tr}\left(\gamma^{-1}\frac{\partial \dot\gamma}{\partial \dot\gamma_{ij}}\gamma^{-1}\dot\gamma + \gamma^{-1}\dot\gamma\gamma^{-1}\frac{\partial \dot\gamma}{\partial \dot\gamma_{ij}}\right)$$

$$= \operatorname{tr}\left(\gamma^{-1}E_{ij}\gamma^{-1}\dot{\gamma} + \gamma^{-1}\dot{\gamma}\gamma^{-1}E_{ij}\right)$$

But $\operatorname{tr}\left((\gamma^{-1}\dot{\gamma})\cdot(\gamma^{-1}E_{ij})\right) = \operatorname{tr}\left((\gamma^{-1}E_{ij})\cdot(\gamma^{-1}\dot{\gamma})\right)$, and thus

$$\frac{\partial\mathcal{E}}{\partial\dot{\gamma}_{ij}} = 2\operatorname{tr}\left(\gamma^{-1}E_{ij}\gamma^{-1}\dot{\gamma}\right)$$

Consequently,

$$\frac{d}{dt}\frac{\partial\mathcal{E}}{\partial\dot{\gamma}_{ij}} = 2\operatorname{tr}\left(\frac{d\gamma^{-1}}{dt}E_{ij}\gamma^{-1}\dot{\gamma} + \gamma^{-1}\frac{dE_{ij}}{dt}\gamma^{-1}\dot{\gamma} + \gamma^{-1}E_{ij}\frac{d\gamma^{-1}}{dt}\dot{\gamma} + \gamma^{-1}E_{ij}\gamma^{-1}\frac{d\dot{\gamma}}{dt}\right)$$

Once again,

$$\frac{d\gamma^{-1}}{dt} = -\gamma^{-1}\dot{\gamma}\gamma^{-1}$$

Thus

$$\frac{d}{dt}\frac{\partial\mathcal{E}}{\partial\dot{\gamma}_{ij}} = 2\operatorname{tr}\left(-\gamma^{-1}\dot{\gamma}\gamma^{-1}E_{ij}\gamma^{-1}\dot{\gamma} - \gamma^{-1}E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma} + \gamma^{-1}E_{ij}\gamma^{-1}\ddot{\gamma}\right)$$

$$= 2\operatorname{tr}\left(-E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} + E_{ij}\gamma^{-1}\ddot{\gamma}\gamma^{-1}\right) = 2\operatorname{tr}\left(E_{ij}(\gamma^{-1}\ddot{\gamma}\gamma^{-1} - 2\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1})\right)$$

Thus, applying the Euler-Lagrange equations (Eq. (1.3)) yield:

$$-2\operatorname{tr}\left(E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1}\right) = 2\operatorname{tr}\left(E_{ij}(\gamma^{-1}\ddot{\gamma}\gamma^{-1} - 2\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1})\right)$$

$$\implies \operatorname{tr}\left(E_{ij}\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1}\right) = \operatorname{tr}\left(E_{ij}(-\gamma^{-1}\ddot{\gamma}\gamma^{-1} + 2\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1})\right) \implies \operatorname{tr}\left(E_{ij}(\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - \gamma^{-1}\ddot{\gamma}\gamma^{-1})\right) = 0$$

But

$$\operatorname{tr}\left(E_{ij}(\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - \gamma^{-1}\ddot{\gamma}\gamma^{-1})\right) = \langle E_{ij}, \gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - \gamma^{-1}\ddot{\gamma}\gamma^{-1}\rangle_{\text{Frob}}$$

Since $\{E_{ij}\}$'s span $\mathcal{S}_k$,

$$\langle S, \gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - \gamma^{-1}\ddot{\gamma}\gamma^{-1}\rangle_{\text{Frob}} = 0$$

for all $S \in \mathcal{S}_k$, and consequently,

$$\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} - \gamma^{-1}\ddot{\gamma}\gamma^{-1} = 0 \implies \ddot{\gamma}\gamma^{-1} - \dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} = 0 \implies \frac{d(\dot{\gamma}\gamma^{-1})}{dt} = 0$$

Consequently,

$$\dot{\gamma}\gamma^{-1} = C \implies \dot{\gamma} = C\gamma$$

This is a matrix-valued differential equation with the solution $\gamma(t) = \exp(tC)\gamma(0)$. Now, suppose we want to find geodesic(s) joining matrices $P, Q \in \mathbb{S}^k_{++}$. Then $\gamma(0) = P$. We still need to find C such that $\gamma(1) = Q$. To that extent, we perform a 'diagonalization' trick, where we write $C = P^{1/2}SP^{-1/2}$. [1] Note that $C^\ell = P^{1/2}S^\ell P^{-1/2}$, for any $\ell \in \mathbb{N}_0$. Then,

$$\gamma(t) = \exp(tC)P = \sum_{\ell=0}^{\infty}\frac{t^\ell}{\ell!}C^\ell P = \sum_{\ell=0}^{\infty}\frac{t^\ell}{\ell!}P^{1/2}S^\ell P^{-1/2}P = \sum_{\ell=0}^{\infty}\frac{t^\ell}{\ell!}P^{1/2}S^\ell P^{1/2} = P^{1/2}\sum_{\ell=0}^{\infty}\frac{(tS)^\ell}{\ell!}P^{1/2} = P^{1/2}\exp(tS)P^{1/2}$$

Thus

$$\exp(S) = P^{-1/2}QP^{-1/2}$$

Now, since $S \in \mathbb{S}^k_{++}$, $S = UD'U^\mathsf{T}$ for some diagonal matrix D'. Then

$$\exp(S) = \sum_{\ell=0}^{\infty}\frac{S^\ell}{\ell!} = \sum_{\ell=0}^{\infty}U\frac{D'^\ell}{\ell!}U^\mathsf{T} = U\exp(D')U^\mathsf{T}$$

[1] Recall that $P \in \mathbb{S}^k_{++}$, and hence $P = ODO^\mathsf{T}$ is diagonalizable, with all eigenvalues being strictly positive. Consequently, for any $\alpha \in \mathbb{R}$, $P^\alpha := OD^\alpha O^\mathsf{T}$, $D^\alpha := \operatorname{diag}(\lambda_1^\alpha, \ldots, \lambda_k^\alpha)$, where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$

But it is easy to see that $\exp(D') = \operatorname{diag}(e^{d_1}, \ldots, e^{d_k})$, where $D' = \operatorname{diag}(d_1, \ldots, d_k)$. Similarly, $\exp(\alpha S) = U \exp(\alpha D') U^{\mathsf{T}} = (\exp(S))^{\alpha}$. Thus,

$$\gamma(t) = P^{1/2} \exp(tS) P^{1/2} = P^{1/2} (\exp(S))^t P^{1/2} = P^{1/2} (P^{-1/2} Q P^{-1/2})^t P^{1/2}$$

Consequently, on $\mathbb{S}^k_{++}$, a geodesic between any two points P, Q is parametrized as $P^{1/2}(P^{-1/2}QP^{-1/2})^t P^{1/2}$.

§2. Geodesic Convexity

We now define the very important notion of geodesic convexity.

Definition 2.1 (Geodesic Convexity)**.** A subset $S \subset M$ of a Riemannian manifold M is geodesically convex if for all $x, y \in S$, there exists a geodesic $\gamma : [0,1] \mapsto M, \gamma(0) = x, \gamma(1) = y$, such that $\gamma([0,1]) \subset S$.

Remark. A few remarks are in order:

1. $\varnothing$ and singleton sets are vacuously geodesically convex.
2. If M is a Riemannian manifold such that there exists a geodesic between any two points, then M is geodesically convex w.r.t. itself.
3. For the Euclidean manifold, geodesically convex sets are convex.

The definition of geodesically convex functions follows almost immediately.

Definition 2.2 (Geodesic Convexity of functions)**.** Let S be a geodesically convex subset of some ambient manifold M. A function $f : S \mapsto \mathbb{R}$ (not necessarily continuous/smooth) is called geodesically (strictly) convex if $f \circ \gamma : [0,1] \mapsto \mathbb{R}$ is (strictly) convex for all geodesics $\gamma : [0,1] \mapsto M$ such that $\gamma(0) \neq \gamma(1)$ and $\gamma([0,1]) \subset S$. In other words, for any geodesic γ, we have

$$f(\gamma(t)) \leq (1-t)f(\gamma(0)) + tf(\gamma(1)), t \in [0,1]$$

if f is geodesically convex, and if f is *strictly* geodesically convex, then

$$f(\gamma(t)) < (1-t)f(\gamma(0)) + tf(\gamma(1)), t \in (0,1)$$

Remark. If f, g are geodesically convex, μ-strongly convex, or strictly convex, then $\lambda f + (1-\lambda)g$ is also geodesically convex, μ-strongly convex, or strictly convex, respectively, for any $\lambda \in (0,1)$.

We now define some related notions in analogy with usual Euclidean convexity.

Definition 2.3 (Geodesic Strong Convexity)**.** Let S be a geodesically convex subset of some ambient manifold M. A function $f : S \mapsto \mathbb{R}$ is called geodesically μ-strongly convex if for all geodesics $\gamma : [0,1] \mapsto M$ with $\gamma([0,1]) \subset S$, we have:

$$f(\gamma(t)) \leq (1-t)f(\gamma(0)) + tf(\gamma(1)) - \frac{t(1-t)\mu}{2}L(\gamma)^2$$

where $L(\gamma)$ is the length of γ (recall Eq. (1.1)). Equivalently, f is geodesically μ-strongly convex if $f \circ \gamma : [0,1] \mapsto \mathbb{R}$ is $\mu L(\gamma)$-strongly convex.

Definition 2.4 (Concavity and Linearity)**.** Let S be a geodesically convex subset of some ambient manifold M. A function $f : S \mapsto \mathbb{R}$ is called geodesically concave if $-f$ is geodesically convex, and f is called geodesically linear if f is both geodesically convex and concave.

2.1. Properties of Convex Functions

Since geodesic convexity has been defined by pulling back along geodesics and then imposing usual convexity, many of the usual properties of convexity transfer through.

Lemma 2.1 (Local Minimizers are Global Minimizers)**.** If $f : S \mapsto \mathbb{R}$ is geodesically convex, then any local minimizer is a global minimizer.

Proof. Let x be a local minimizer, and assume for the sake of contradiction we have some $y \in S$ such that $f(y) < f(x)$. Since S is geodesically convex, there exists a geodesic γ connecting x to y. Then for all $t \in (0, 1]$

$$f(\gamma(t)) \leq (1-t)f(x) + tf(y) = f(x) + t(f(y) - f(x)) < f(x)$$

Since γ is smooth, $\lim_{t \searrow 0} \gamma(t) = x$, and consequently, for any neighborhood U of x, there exists $t_U > 0$ such that $\gamma(t_U) \in U$, which contradicts the fact that x is a local minimizer, since in any neighborhood of x we are able to find points x' such that $f(x') < f(x)$. ■

This result immediately yields many useful corollaries.

Corollary 2.2. If $f : S \mapsto \mathbb{R}$ is geodesically strictly convex, then it has at most one local minimizer, which, by the above lemma, must also be a global minimizer.

Proof. Assume for the sake of contradiction that f has two global minimizers x, y. Then we must have $f(x) = f(y)$. Let γ be a geodesic connecting x, y. Then for any $t \in (0, 1)$,

$$f(\gamma(t)) < (1-t)f(x) + tf(y) = f(x)$$

which leads to a contradiction. ■

Remark. Note that a global minimizer may not always exist: For example, $f : [1, \infty) \mapsto \mathbb{R}$, $f(x) := 1/x$ is a strictly convex function with no global minimizer.

Corollary 2.3. Let M be a smooth manifold, and let $f : M \mapsto \mathbb{R}$ be a function such that f has a local minimizer which is not a global minimizer, i.e. there is a point $p \in M$, and a neighborhood U_p of p, such that

$$\inf_{x \in M} f(x) < f(p) = \inf_{x \in U_p} f(x)$$

Then there does not exist any metric tensor on M such that f is geodesically convex w.r.t that metric tensor.

Some more of the usual stuff also holds:

Lemma 2.4 (Sublevel Sets are Convex)**.** Let $\{f_i\}_{i \in \mathcal{I}}$ be an arbitrary collection of geodesically convex functions $f_i : S \mapsto \mathbb{R}$. Consider a collection of real numbers $\{\alpha_i\}_{i \in \mathcal{I}}$. Define the *sublevel sets*

$$S_i := f_i^{-1}\left((-\infty, \alpha_i]\right) = \{x \in S : f_i(x) \leq \alpha_i\}$$

Then $S' := \bigcap_{i \in \mathcal{I}} S_i$ is geodesically convex. In particular, if f is a geodesically convex function, then all sublevel sets of f are geodesically convex.

Proof. Empty sets are geodesically convex, so assume S' is non-empty. Pick any $x, y \in S'$, and let $\gamma : [0,1] \mapsto M, \gamma([0,1]) \subset S, \gamma(0) = x, \gamma(1) = y$ be a geodesic in S, connecting x and y. Now, for any $i \in \mathcal{I}$,

$$f_i(\gamma(t)) \leq (1-t)f_i(x) + tf_i(y) \leq (1-t)\alpha_i + t\alpha_i = \alpha_i$$

Thus $\gamma([0,1]) \subset S_i$ for all i, and thus $\gamma([0,1]) \subset S'$, as desired. ■

We also state without proof the following results. See [UDR77].

Lemma 2.5. If $f : S \mapsto \mathbb{R}$ is geodesically convex, then f is continuous on the interior of S.

Remark. Recall that in a topological space, the interior of a set A is defined to be the union of all open sets it contains.

Corollary 2.6. If M is a connected, compact Riemannian manifold, and if $f : M \mapsto \mathbb{R}$ is geodesically convex, then f is constant.

Consequently, on compact manifolds (like $S^n, \mathbb{T}^n, O(n)$), geodesic convexity is interesting to study only on proper subsets. We now also seek to formalize the idea that convex functions are maximized on the boundary. To do that, we first define a special notion of interior:

Definition 2.5 (Relative Interior)**.** Let S be geodesically convex in the Riemannian manifold M. A point $x \in S$ belongs to the relative interior of S only if for every $y \in S$, and every geodesic $\gamma : [0,1] \mapsto M, \gamma(0) = x, \gamma(1) = y, \gamma([0,1]) \subset S$, there exists a $\varepsilon > 0$ and a geodesic $\nu : [-\varepsilon, 1] \mapsto M$ such that $\nu([-\varepsilon, 1]) \subset S$, and $\nu|_{[0,1]} = \gamma$.
In words, every geodesic connecting x to y within S can be extended to a geodesic in S, beyond x.

Theorem 2.7. Let $f : S \mapsto \mathbb{R}$ be geodesically convex. If f attains its maximum at a point x in the relative interior of S, then f is a constant function.

Proof. Let $y \in S$ be arbitrary. Let $\gamma : [0,1] \mapsto M$ be a geodesic of M connecting x to y within S. Extend γ to ν, and let $z = \nu(-\varepsilon)$. Since f is geodesically convex, $f \circ \nu$ is convex, and we have

$$f(x) \leq \frac{1}{1+\varepsilon}f(z) + \frac{\varepsilon}{1+\varepsilon}f(y) \implies (1+\varepsilon)f(x) \leq f(z) + \varepsilon f(y)$$

Now, since x is a maximizer, $f(x) \geq f(z)$, and consequently

$$(1+\varepsilon)f(x) \leq f(z) + \varepsilon f(y) \leq f(x) + \varepsilon f(y) \implies \varepsilon f(x) \leq \varepsilon f(y)$$

But x is a maximizer, and thus $f(x) \geq f(y)$, implying that $f(x) = f(y)$, as desired. ■

2.2. Other definitions of convexity

Let S be a geodesically convex subset of M. Let $x, y \in S$. Note that geodesic convexity only requires *a* geodesic between x and y to be present in S: In particular, we don't require all geodesics between x and y to be present in S, and neither do we require the geodesic in S to be length minimizing (whenever that makes sense).
In a way, this permissive definition is well-suited for various purposes, since many sets can be included within this

umbrella of definitions. However, it also leads to pathologies as follows: Let $S^1 := \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ be a Riemannian submanifold of $\mathbb{R}^2$, i.e. S^1 inherits both its smoothness and its metric tensor, from $\mathbb{R}^2$. Then $A := \{(x_1, x_2) \in S^1 : x_1 \geq 0\}, B := \{(x_1, x_2) \in S^1 : x_1 \leq 0\}$ are both geodesically convex subsets of S^1, yet $A \cap B = \{(0,1), (0,-1)\}$ is not geodesically convex.
To avoid pathologies like this, we define a stronger notion of *total convexity*:

Definition 2.6 (Geodesic Total Convexity)**.** Let M be a manifold such that there is a geodesic between any two points in M. A subset $S \subset M$ is called geodesically totally convex if for any $x, y \in S$, all geodesics between x and y lie in S, i.e. if $\gamma : [0,1] \mapsto M$ is any geodesic such that $\gamma(0) = x, \gamma(1) = y$, then $\gamma([0,1]) \subset S$.

However, total convexity is not the only way to generalize the notion of geodesic convexity. Indeed,

Definition 2.7 (Geodesic Strong Convexity)**.** Let M be a manifold. A subset $S \subset M$ is called geodesically strongly convex, if, for any $x, y \in S$, there exists a unique length-minimizing geodesic $\gamma_{\text{lmin}} : [0,1] \mapsto M$ such that $\gamma(0) = x, \gamma(1) = y$. Furthermore, we also demand that $\gamma_{\text{lmin}}([0,1]) \subset S$.

It is clear that if S is geodesically totally convex, then it is geodesically convex. Similarly, if S is geodesically strongly convex, then it is geodesically convex. Furthermore, unlike geodesic convexity, geodesically total convex sets are closed under intersections, as are geodesically strong convex sets.
However, total convexity and strong convexity have no implications within each other. To highlight this point, consider the following example:

Example. Consider the manifold $S^n := \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : x_1^2 + \cdots + x_{n+1}^2 = 1\}$ embedded in $\mathbb{R}^{n+1}$. Consider the spherical cap

$$S_\alpha := \{(x_1, \ldots, x_{n+1}) \in S^n : x_1 \geq \alpha\}$$

Now, recall that the geodesics on S^n were obtained through the 'great circles', i.e. let $x, y \in S^n$ be such that x, y are not diametrically opposite, and let Γ be the unique equator of S^n containing both x, y. Then, x, y split Γ into 2 arcs, both of which are geodesics. However, the shorter arc is the unique length minimizing geodesic between x, y. If x, y are diametrically opposite, then there are infinitely many equators containing both x, y, and each such equator gives rise to 2 geodesics, and all these geodesics have the same length.
From this description, it is easy to see that S_α is geodesically convex for all $\alpha \in \mathbb{R}$ (note that if $\alpha \leq -1$, then $S_\alpha = S^n$, and if $\alpha > 1$, then $S_\alpha = \varnothing$).
However, S_α is geodesically strongly convex if and only if $\alpha > 0$. Furthermore, S_α is geodesically totally convex if and only if $S_\alpha = \varnothing, S^n$, i.e. $\alpha \notin (-1, 1]$. Clearly, strong convexity and total convexity don't imply each other.

2.3. Differentiable Convex Functions

We eventually hope to build a theory of optimization of geodesically convex functions. For such purposes, having some differentiability helps.

Theorem 2.8 (Differentiability and Convexity)**.** Let S be a geodesically convex set, and let $f : M \mapsto \mathbb{R}$ be differentiable in a neighborhood of S. Then $f|_S$ is geodesically convex if and only if for every geodesic $\gamma : [0,1] \mapsto M, \gamma([0,1]) \subset S$, we have:

$$f(\gamma(t)) \geq f(\gamma(0)) + t\langle \nabla f, \gamma'(0) \rangle_{g(\gamma(0))}, \forall t \in [0,1]$$

$f|_S$ is geodesically μ-strongly convex if and only if

$$f(\gamma(t)) \geq f(\gamma(0)) + t\langle \nabla f, \gamma'(0) \rangle_{g(\gamma(0))} + \frac{\mu t^2}{2} L(\gamma)^2, \forall t \in [0,1] \tag{2.1}$$

$f|_S$ is geodesically strictly convex if and only if whenever $\gamma'(0) \neq 0$, we have

$$f(\gamma(t)) > f(\gamma(0)) + t\langle \nabla f, \gamma'(0)\rangle_{g(\gamma(0))}, \forall t \in (0,1]$$

Proof. Suppose $f \circ \gamma$ is differentiable. Then it is convex if and only if for all $s, t \in [0,1]$, we have:

$$f(\gamma(t)) \geq f(\gamma(s)) + (t-s)(f \circ \gamma)'(s)$$

Now,

$$(f \circ \gamma)'(s) = Df(\gamma(s)) \cdot \gamma'(s)$$

But $Df(\gamma(s))$ will act upon $\gamma'(s)$ according to the Riemannian metric tensor, and thus

$$D(f(\gamma(s))) \cdot \gamma'(s) = \langle (\nabla f)(\gamma(s)), \gamma'(s)\rangle_{g(\gamma(s))}$$

where g is the metric tensor. Putting $s = 0$ gets us what we want. The proof for μ-strong convexity is similar.
In the other direction, if we have

$$f(\gamma(t)) \geq f(\gamma(0)) + t\langle \nabla f, \gamma'(0)\rangle_{g(\gamma(0))}, \forall t \in [0,1]$$

for all geodesics γ contained in S, then we first rewrite it as

$$f(\gamma(t)) \geq f(\gamma(0)) + t(f \circ \gamma)'(0), \forall t \in [0,1]$$

Now, let μ be any geodesic, and let $s, t \in [0,1]$ be arbitrary, with $s \leq t$. Set $\nu(x) := \mu(s+(t-s)x)$. Since sub-segments of geodesics are also geodesics, ν is a geodesic. Consequently, we have:

$$f(\nu(1)) \geq f(\nu(0)) + (f \circ \nu)'(0)$$

But $\nu(1) = \mu(t), \nu(0) = \mu(s)$, and thus

$$f(\mu(t)) \geq f(\mu(s)) + (f \circ \nu)'(0)$$

Now, consider the map $[0,1] \ni x \mapsto s + (t-s)x =: \tau(x) \in \mathbb{R}$. Then $\nu = \mu \circ \tau$, and thus

$$(f \circ \nu)'(0) = (f \circ \mu \circ \tau)'(0) = (f \circ \mu)'(\tau(0)) \cdot \tau'(0) = (f \circ \mu)'(s) \cdot (t-s)$$

Consequently,

$$f(\mu(t)) \geq f(\mu(s)) + (t-s) \cdot (f \circ \mu)'(s)$$

Thus, $f \circ \mu$ is convex. Since μ was an arbitrary geodesic, f is geodesically convex. The proofs for strong convexity and strict convexity follow similarly. ∎

Corollary 2.9. If $f : S \mapsto \mathbb{R}$ is differentiable and geodesically convex, where S is open and geodesically convex, then x is a global minimizer of f if and only if $(\nabla f)(x) = 0$.

Proof. If $(\nabla f)(x) = 0$, then $f(\gamma(t)) \geq f(x)$ for any $t \in [0,1]$. Since f is defined on a geodesically convex set, for any point $x' \in S$, we can find a geodesic γ with $\gamma(1) = x'$, thus yielding $f(x') \leq f(x)$ (using Theorem 2.8), i.e. x is a global minimizer. Note that we didn't need to use the openness of S for this implication.
Conversely, let x be a global minimizer of f. Then $f(x') \geq f(x)$ for all $x' \in S$. Now, let $\xi : [-1,1] \mapsto S$ be any smooth map [2] such that $\xi(0) = x$. Assume for the sake of contradiction that $(f \circ \xi)'(0) < 0$. Since $(f \circ \xi)'(0) < 0$, there exists $\delta > 0$ such that $(f \circ \xi)'(\alpha) < 0$ for all $\alpha \in [0, \delta]$. Consequently, for any $\alpha \in (0, \delta]$, we have:

$$f(\xi(\alpha)) = f(\xi(0)) + \int_0^{\alpha} (f \circ \xi)'(s) ds < f(\xi(0)) = f(x)$$

which is a contradiction.
We now claim that $(f \circ \xi)'(0)$ is actually 0: Otherwise, set $\nu(\alpha) := \xi(-\alpha)$, and we get that $(f \circ \nu)'(0) = -(f \circ \xi)'(0) < 0$. Since $(f \circ \xi)'(0) = (\nabla f)(x) \cdot \xi'(0) = 0$ for all smooth curves $\xi : [-1,1] \mapsto S$, we must have $(\nabla f)(x) = 0$, as desired. ∎

[2] since S is an open subset of M, we can consider it to be an embedded submanifold of M, based on which smoothness of maps can be defined

We now also give characterizations based on second-order derivatives, without proof (the proof is essentially identical to that in standard convex analysis, except for the fact that we have to pull f back along the geodesics):

Theorem 2.10. Let $f : S \mapsto \mathbb{R}$ be twice-differentiable, and assume S is open and geodesically convex. Then f is:

1. Geodesically convex if and only if $\text{Hess } f(x) \succeq 0$.
2. Geodesically μ-strongly convex if and only if $\text{Hess } f(x) \succeq \mu \cdot \text{Id}$.
3. Geodesically strictly convex if $\text{Hess } f(x) \succ 0$. Note that this is just an 'if' condition, not an 'if and only if' condition.

In all of the above conditions, it is assumed that $x \in S$ is arbitrary.

2.4. Some Examples

Now that we have studied the properties of geodesic convex functions in some detail, let's see a few examples.

Lemma 2.11 (Geodesic Linearity of the log-barrier function)**.** The map $\mathbb{R}^n_{>0} \ni x \mapsto \langle 1, \ln(x) \rangle \in \mathbb{R}$ is geodesically linear on the manifold $\mathbb{R}^n_{>0}$ as defined in the first chapter.

Proof. Note that the geodesics on $\mathbb{R}^n_{>0}$ are of the form $\exp(\alpha t + \beta), \alpha, \beta \in \mathbb{R}^n$ (the exponential function is evaluated coordinate-wise). Thus, the restriction of $\langle 1, \ln(x) \rangle$ on a geodesic yields $\langle 1, \alpha \rangle t + \langle 1, \beta \rangle$, which is linear, as desired. ■

Remark. At some level, this result shouldn't be so surprising: The manifold $\mathbb{R}^n_{>0}$ (with the log-barrier metric tensor) is a Hessian manifold, with the Hessian being generated from the log-barrier function. Consequently, the log-barrier function will be geodesically linear on this manifold. We shall see another example of this phenomenon with the manifold $\mathbb{S}^n_{++}$ soon.

Lemma 2.12 (Geodesic Convexity of Polynomials with positive coefficients)**.** Let $p(x_1, \dots, x_n)$ be a multivariate polynomial with positive coefficients. Then p is geodesically convex on $\mathbb{R}^n_{>0}$.

Proof. If we can show that the monomial $x^\lambda := \prod_{i=1}^n x_i^{\lambda_i}, \lambda \in \mathbb{N}_0^n$ is geodesically convex on $\mathbb{R}^n_{>0}$, then we're done since p is a positive linear combination of such monomials, and thus the Hessian of $p(\gamma(t))$ will also be a positive linear combination of PSD matrices, which will be PSD.
Now, x^λ evaluated on a geodesic yields $\exp(\langle \lambda, \alpha \rangle t + \langle \lambda, \beta \rangle)$, which is convex, as desired. ■

In fact, we will now show that not only are multivariate polynomials geodesically convex, but they are geodesically log-convex, which is a much more powerful property.

Lemma 2.13 (Geodesic Log-Convexity of Polynomials with positive coefficients)**.** Let $p(x_1, \dots, x_n)$ be a multivariate polynomial with positive coefficients. Then p is geodesically log-convex on $\mathbb{R}^n_{>0}$.

Proof. Fix a geodesic $\gamma(t) := \exp(\alpha t + \beta)$. Let $p(x) = \sum_{\lambda \in \mathbb{N}_0^n} c_\lambda x^\lambda$, where $c_\lambda \geq 0$. Then

$$p(\gamma(t)) = \sum_{\lambda \in \mathbb{N}_0^n} c_\lambda \exp(\langle \lambda, \alpha \rangle t + \langle \lambda, \beta \rangle)$$

Then

$$\frac{d\ln p(\gamma(t))}{dt} = \frac{\sum_{\lambda\in\mathbb{N}_0^n} c_\lambda\langle\lambda,\alpha\rangle\exp(\langle\lambda,\alpha\rangle t + \langle\lambda,\beta\rangle)}{\sum_{\lambda\in\mathbb{N}_0^n} c_\lambda\exp(\langle\lambda,\alpha\rangle t + \langle\lambda,\beta\rangle)}$$

$$\frac{d^2\ln p(\gamma(t))}{dt^2} = \frac{\sum_{\lambda,\lambda'\in\mathbb{N}_0^n}(c_\lambda\langle\lambda,\alpha\rangle - c_{\lambda'}\langle\lambda',\alpha\rangle)^2\exp(\langle\lambda,\alpha\rangle t + \langle\lambda,\beta\rangle)\exp(\langle\lambda',\alpha\rangle t + \langle\lambda',\beta\rangle)}{\left(\sum_{\lambda\in\mathbb{N}_0^n} c_\lambda\exp(\langle\lambda,\alpha\rangle t + \langle\lambda,\beta\rangle)\right)^2}$$

Both the numerator and denominator are non-negative, as desired. ■

We shall now give analogs of all the above examples for the manifold $\mathbb{S}_{++}^n$. There is indeed an analogy between $\mathbb{R}_{>0}^n$ and $\mathbb{S}_{++}^n$ in the sense that they are both Hessian manifolds, that too of "log-barrier functions" ($\sum\ln(x_i)$ for $\mathbb{R}_{>0}^n$, $\ln\det(X)$ for $\mathbb{S}_{++}^n$).

Lemma 2.14 (Geodesic Linearity of the log-det map)**.** The map $\mathbb{S}_{++}^n \ni X \mapsto \ln\det(X) \in \mathbb{R}$ is geodesically linear on the manifold $\mathbb{S}_{++}^n$ as defined in the first chapter.

Proof. Let $X, Y \in \mathbb{S}_{++}^n$. Then there is a unique geodesic joining X, Y, which is given by

$$\gamma(t) := X^{1/2}(X^{-1/2}YX^{-1/2})^tX^{1/2}$$

Thus

$$\ln\det\gamma(t) = \ln\det(X^{1/2}(X^{-1/2}YX^{-1/2})^tX^{1/2}) = \ln\det(X)+t(\ln\det(Y)-\ln\det(X)) = (1-t)\ln\det(X)+t\ln\det(Y)$$

Thus $\ln\det(\cdot)$ is a geodesically linear function. ■

Lemma 2.15 (Geodesic Convexity of Strictly Positive Linear Operators)**.** Let $T : \mathcal{S}_n \mapsto \mathcal{S}_m$ be a linear map such that $T(\mathbb{S}_{++}^n) \subseteq \mathbb{S}_{++}^m$. Such maps are also called strictly positive linear operators. Then $\mathbb{S}_{++}^n \ni X \mapsto T(X) \in \mathbb{S}_{++}^m$ is a geodesically convex map w.r.t the order $\preceq$ on $\mathbb{S}_{++}^m$, i.e. for any geodesic $\gamma : [0,1] \mapsto \mathbb{S}_{++}^n$, and any $t \in [0,1]$, we have

$$T(\gamma(t)) \preceq (1-t)T(\gamma(0)) + tT(\gamma(1))$$

Proof. It is easy to see that any linear map from $\mathbb{S}_{++}^n$ to $\mathbb{S}_{++}^m$ has to be of the form:

$$T(X) := \sum_{i\in[d]} A_iXB_i$$

for some matrices $A_i \in \mathbb{R}^{m\times n}, B_i \in \mathbb{R}^{n\times m}$. Let $\gamma(t) := P^{1/2}\exp(tQ)P^{1/2}$ be an arbitrary geodesic in $\mathbb{S}_{++}^n$. Then

$$T(\gamma(t)) = \sum_{i\in[d]} A_iP^{1/2}\exp(tQ)P^{1/2}B_i$$

$$\frac{dT(\gamma(t))}{dt} = \sum_{i\in[d]} A_iP^{1/2}Q\exp(tQ)P^{1/2}B_i = T(P^{1/2}Q\exp(tQ)P^{1/2}) \tag{2.2}$$

where we use the fact that (see Eq. (A.3))

$$\frac{d\exp(tQ)}{dt} = Q\exp(tQ)$$

Now, also note that

$$\frac{d^2 T(\gamma(t))}{dt^2} = \sum_{i\in[d]} A_i P^{1/2} Q^2 \exp(tQ) P^{1/2} B_i = T(P^{1/2} Q^2 \exp(tQ) P^{1/2}) \tag{2.3}$$

Since Q and $\exp(tQ)$ commute,

$$T(P^{1/2} Q^2 \exp(tQ) P^{1/2}) = T(P^{1/2} Q \exp(tQ) Q P^{1/2})$$

Now, we claim that $P^{1/2} Q \exp(tQ) Q P^{1/2}$ is positive semi-definite (note that this is not obvious, since $Q \in \mathcal{S}_n$): Note that since $P^{1/2}$ is positive definite, it suffices to show that $Q \exp(tQ) Q$ is positive semi-definite. To that extent, it is easy to see that if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of Q, then the eigenvalues of $Q \exp(tQ) Q$ are $\lambda_1^2 e^{t\lambda_1}, \ldots, \lambda_n^2 e^{t\lambda_n}$, which are obviously non-negative.
Now, since T is continuous, [3]

$$T\left(\overline{\mathbb{S}_{++}^n}\right) \subseteq \overline{T(\mathbb{S}_{++}^n)} \subseteq \overline{\mathbb{S}_{++}^m}$$

Thus T maps positive semi-definite matrices to positive semi-definite matrices, and thus $T(P^{1/2} Q \exp(tQ) Q P^{1/2})$ is positive semi-definite, whence we're done by Theorem 2.10. ∎

Lemma 2.16 (Geodesic Convexity of log-det of Strictly Positive Operators)**.** Let $T : \mathcal{S}_n \mapsto \mathcal{S}_m$ be a linear map such that $T(\mathbb{S}_{++}^n) \subseteq \mathbb{S}_{++}^m$. Then $\mathbb{S}_{++}^n \ni X \mapsto \ln\det(T(X)) \in \mathbb{R}$ is a geodesically convex map.

Proof. Let $\gamma(t) := P^{1/2} \exp(tQ) P^{1/2}$ be an arbitrary geodesic in $\mathbb{S}_{++}^n$. By Theorem 2.10, if we can show that

$$\frac{d^2 \ln\det(T(\gamma(t)))}{dt^2} \geq 0$$

then we'd be done. Since $\mathbb{S}_{++}^n$ is geodesically complete, and we can start a geodesic from any point with any velocity, it in fact suffices to show that

$$\left.\frac{d^2 \ln\det(T(\gamma(t)))}{dt^2}\right|_{t=0} \geq 0$$

Now, we first recall *Jacobi's formula*:

$$\frac{d\det(A(t))}{dt} = \det(A(t)) \cdot \operatorname{tr}\left(A(t)^{-1} \cdot \dot{A}(t)\right)$$

Then

$$\frac{d\ln\det(T(\gamma(t)))}{dt} = \frac{1}{\det(T(\gamma(t)))} \cdot \det(T(\gamma(t))) \cdot \operatorname{tr}\left(T(\gamma(t))^{-1} \cdot \frac{dT(\gamma(t))}{dt}\right) = \operatorname{tr}\left(T(\gamma(t))^{-1} \cdot \frac{dT(\gamma(t))}{dt}\right)$$

$$\begin{aligned}\frac{d^2 \ln\det(T(\gamma(t)))}{dt^2} &= \frac{d}{dt}\operatorname{tr}\left(T(\gamma(t))^{-1} \cdot \frac{dT(\gamma(t))}{dt}\right) = \operatorname{tr}\left(\frac{dT(\gamma(t))^{-1}}{dt} \cdot \frac{dT(\gamma(t))}{dt} + T(\gamma(t))^{-1} \cdot \frac{d^2T(\gamma(t))}{dt^2}\right) \\ &= \operatorname{tr}\left(-T(\gamma(t))^{-1} \cdot \frac{dT(\gamma(t))}{dt} \cdot T(\gamma(t))^{-1} \cdot \frac{dT(\gamma(t))}{dt} + T(\gamma(t))^{-1} \cdot \frac{d^2T(\gamma(t))}{dt^2}\right) \\ &= \operatorname{tr}\left(T(\gamma(t))^{-1} \cdot \left(\frac{d^2T(\gamma(t))}{dt^2} - \frac{dT(\gamma(t))}{dt} \cdot T(\gamma(t))^{-1} \cdot \frac{dT(\gamma(t))}{dt}\right)\right)\end{aligned}$$

[3] if f is a continuous map, then $f(\overline{U}) \subseteq \overline{f(U)}$

Thus, the evaluation of the above second derivative, at zero, yields, by Eq. (2.2), Eq. (2.3),

$$\operatorname{tr}\left(T(P)^{-1}\cdot\left(T(P^{1/2}Q^2P^{1/2})-T(P^{1/2}QP^{1/2})\cdot T(P)^{-1}\cdot T(P^{1/2}QP^{1/2})\right)\right)$$

Since $\operatorname{tr}(\cdot)$ is the sum of eigenvalues, and since $T(P)\succ 0$, it suffices to show that

$$T(P^{1/2}Q^2P^{1/2})\succeq T(P^{1/2}QP^{1/2})\cdot T(P)^{-1}\cdot T(P^{1/2}QP^{1/2})$$

More generally, if $\mathcal{S}_n\ni X\mapsto T'(X):=T(P)^{-1/2}T(P^{1/2}XP^{1/2})T(P)^{-1/2}$, and if we can show that $T'(X^2)\succeq T'(X)^2$, then the above result follows by putting $X=Q$.
Now, a classic result from matrix algebra says that if A,B,C,D are matrices such that A,D are square matrices, D is invertible, and the expression $A-BD^{-1}C$ is well-defined, then $A\succeq BD^{-1}C$ if and only if

$$\begin{bmatrix}A & B\\ C & D\end{bmatrix}\succeq 0$$

Thus, we wish to show that

$$\begin{bmatrix}T'(X^2) & T'(X)\\ T'(X) & I_m\end{bmatrix}\succeq 0$$

After this point, it is just bashing, by substituting $X=\sum_{i=1}^n\lambda_iu_iu_i^\intercal$ (this follows from the spectral theorem). Indeed,

$$X^2=\sum_{i=1}^n\lambda_i^2u_iu_i^\intercal, I_m=T'(I_n)=T'\left(\sum_{i=1}^n u_iu_i^\intercal\right)=\sum_{i=1}^n T'(u_iu_i^\intercal)$$

Then

$$\begin{bmatrix}T'(X^2) & T'(X)\\ T'(X) & I\end{bmatrix}=\sum_{i=1}^n\begin{bmatrix}T'(\lambda_i^2u_iu_i^\intercal) & T'(\lambda_iu_iu_i^\intercal)\\ T'(\lambda_iu_iu_i^\intercal) & T'(u_iu_i^\intercal)\end{bmatrix}$$

Set $U_i:=T'(u_iu_i^\intercal)$. Since $u_iu_i^\intercal$ is PSD, U_i is PSD. Then

$$\begin{bmatrix}T'(\lambda_i^2u_iu_i^\intercal) & T'(\lambda_iu_iu_i^\intercal)\\ T'(\lambda_iu_iu_i^\intercal) & T'(u_iu_i^\intercal)\end{bmatrix}=\begin{bmatrix}\lambda_i^2U_i & \lambda_iU_i\\ \lambda_iU_i & U_i\end{bmatrix}=\begin{bmatrix}\lambda_i^2 & \lambda_i\\ \lambda_i & 1\end{bmatrix}\otimes U_i$$

Since $\begin{bmatrix}\lambda_i^2 & \lambda_i\\ \lambda_i & 1\end{bmatrix}$ is PSD, and U_i is PSD, their tensor product is PSD too. Finally, since $\begin{bmatrix}T'(X^2) & T'(X)\\ T'(X) & I\end{bmatrix}$ is a sum of PSD matrices, it is PSD too, as desired. ■

2.5. Riemannian Gradient Descent

Before we get to Gradient Descent on Riemannian manifolds, we first recall the exponential map. The exponential map helps us rephrase many things more elegantly: For example, instead of saying that there is a geodesic connecting x and y, we can say that there exists some $v\in T_xM$ such that $\exp_x(v)=y$.
Note that we can now rephrase Eq. (2.1) as:

$$f(\exp_x(tv))\geq f(x)+t\langle\nabla f(x),v\rangle_x+\frac{\mu t^2}{2}\cdot\|v\|_x^2 \tag{2.4}$$

where we use the fact that $L(\gamma_{x,v}|_{[0,1]})=\|v\|_x=:\sqrt{\langle v,v\rangle_x}$ (see Definition A.2).
The way we use this inequality for optimization is by topping it off with an upper bound: Indeed, if ∇f is L-Lipschitz continuous, then

$$f(\exp_x(tv))\leq f(x)+t\langle\nabla f(x),v\rangle_x+\frac{Lt^2}{2}\cdot\|v\|_x^2 \tag{2.5}$$

We shall now slowly build up towards using these things to design Riemannian Gradient Descent. Before, that, we need some lemmata:

Lemma 2.17. Let S be a non-empty, closed, and geodesically convex subset of M, where M is a complete Riemannian manifold. Assume $f : M \mapsto \mathbb{R}$ is differentiable on a neighborhood of S. If $f|_S$ is geodesically μ-strongly convex with $\mu > 0$, then the sublevel sets of $f|_S$ are compact, and $f|_S$ has exactly one global minimizer.

Proof. Let $x_0 \in S$ be arbitrary. We first prove that the sublevel set $S_0 := \{x \in S : f(x) \leq f(x_0)\}$ is compact. Firstly, S_0 is closed, since $S_0 = (f|_S)^{-1}\left((-\infty, f(x_0)]\right)$ is the pre-image of a closed set under a continuous map. Secondly, we claim that S_0 is bounded: Suppose not. Then there is a sequence $x_1, x_2, x_3, \ldots \in S_0$ such that $d(x_0, x_k) \xrightarrow{k\to\infty} \infty$. Since S is geodesically convex, there exist geodesics between x_0 and x_k for all $k \in \mathbb{N}$, and thus let $v_k \in T_{x_0}M$ be such that $\gamma_{x_0,v_k}(1) = x_k$. Then by Eq. (2.4), we have

$$f(x_k) \geq f(x_0) + \langle \nabla f(x_0), v_k \rangle_{x_0} + \frac{\mu}{2}\|v_k\|_{x_0}^2 \tag{2.6}$$

Now, recall that

$$\|v_k\|_{x_0} = L(\gamma_{x,v_k}|_{[0,1]}) \geq d(x_0, x_k)$$

where the last inequality follows since $d(x_0, x_k)$ is the infimum of the lengths of all paths joining x_0, x_k. Thus, since $d(x_k, x_0) \to \infty$, we have that $\|v_k\|_{x_0} \to \infty$ as $k \to \infty$, and consequently, by Eq. (2.6), we have that $f(x_k) \xrightarrow{k\to\infty} \infty$, which is a contradiction, since $x_k \in S_0$, which entails $f(x_k) \leq f(x_0)$. Since S_0 is closed and bounded, and since M is complete (as a metric space), by the Hopf-Rinow theorem (see Theorem A.2), S_0 is compact, as desired.
Now, since S_0 is compact, and since f is continuous over S_0, f attains its minima at $x_* \in S_0$. Now, for any $x \in S$, if $x \notin S_0$, then $f(x) > f(x_0) \geq f(x_*)$. If $x \in S_0$, then $f(x) \geq f(x_*)$. Consequently, x_* is the minimizer of $f|_S$. Since f is strictly convex, by Corollary 2.2, we have that x_* is the unique global minimizer of $f|_S$. ∎

The above lemma can be quantitatively sharpened to obtain estimates about $f(x_*)$, which we shall need later.

Lemma 2.18 (Polyak-Łojasiewicz Inequality)**.** Let S be a non-empty, closed, and geodesically convex subset of M, where M is a complete Riemannian manifold. Assume $f : M \mapsto \mathbb{R}$ is differentiable on a neighborhood of S. If $f|_S$ is geodesically μ-strongly convex with $\mu > 0$, then

$$f(x) - f(x_*) \leq \frac{1}{2\mu}\|\nabla f(x)\|_x^2$$

for all $x \in S$, where x_* is the unique global minimizer of $f|_S$.

Proof. Since $x, x_* \in S$, and since S is geodesically convex, there exists $v_x \in T_xM$ such that $x_* = \exp_x(v_x)$, and $[0,1] \ni t \mapsto \exp_x(tv_x) \in S$. Then by Eq. (2.4), we have

$$f(x_*) = f(\exp_x(v_x)) \geq f(x) + \langle \nabla f(x), v_x \rangle_x + \frac{\mu}{2}\|v_x\|_x^2 \geq f(x) + \inf_{v \in T_xM}\left(\langle \nabla f(x), v\rangle_x + \frac{\mu}{2}\|v\|_x^2\right)$$

Now, if $G = g(x)$ is the evaluation of the Riemannian metric tensor at x, then

$$\langle \nabla f(x), v \rangle_x + \frac{\mu}{2}\|v\|_x^2 = (\nabla f(x))^\mathsf{T} G v + \frac{\mu}{2} v^\mathsf{T} G v$$

This expression is a quadratic in v, and is minimized at $v = -\frac{\nabla f(x)}{\mu}$, at which point the expression evaluates to $-\|\nabla f(x)\|_x^2/(2\mu)$, and thus

$$f(x_*) \geq f(x) - \frac{\|\nabla f(x)\|_x^2}{2\mu}$$

as desired. ∎

We are finally ready to prove results about Riemannian Gradient Descent!

Theorem 2.19 (Riemannian Gradient Descent)**.** Let $f : M \mapsto \mathbb{R}$ be a differentiable geodesically convex function on a complete connected manifold M. Let $x_0 \in M$, and consider the sublevel set $S_0 := \{x \in M : f(x) \leq f(x_0)\}$. Assume f has a L-Lipschitz continuous gradient on a neighborhood of S_0, and suppose $f|_{S_0}$ is geodesically μ-strongly convex. Consider **gradient descent with exponential retraction** and step-size $1/L$ initialized at x_0, i.e.

$$x_{k+1} = \exp_{x_k}\left(-\frac{1}{L}\nabla f(x_k)\right), k \in \mathbb{N}_0$$

By Lemma 2.17, there exists a unique global minimizer of $f|_{S_0}$ which is $x_* \in S_0$, and convergence to x_* is linear, i.e. if we set $\kappa = L/\mu$, then we have that $x_k \in S_0$ for all $k \in \mathbb{N}_0$, and

$$f(x_k) - f(x_*) \leq \left(1 - \frac{1}{\kappa}\right)^k (f(x_0) - f(x_*))$$

$$d(x_k, x_*) \leq \sqrt{\kappa} \cdot \left(1 - \frac{1}{\kappa}\right)^{k/2} d(x_0, x_*)$$

Proof. We argue by induction that $x_k \in S_0$ for all $k \in \mathbb{N}_0$. The base case is clear. Suppose $x_k \in S_0$. Now, consider the curve

$$\tau(t) := \exp_{x_k}(-t\nabla f(x_k))$$

Then by Eq. (2.5),

$$f(\tau(t)) \leq f(x_k) - t\left(1 - t\frac{L}{2}\right)\|\nabla f(x_k)\|^2_{x_k}$$

Since $\tau(1/L) = x_{k+1}$, we have

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2_{x_k} \leq f(x_k) \leq f(x_0)$$

where the last inequality follows since $x_k \in S_0$. Consequently, we also have $x_{k+1} \in S_0$. Furthermore,

$$f(x_{k+1}) - f(x_*) \leq f(x_k) - f(x_*) - \frac{1}{2L}\|\nabla f(x_k)\|^2_{x_k}$$

At the same time, by Lemma 2.18,

$$\|\nabla f(x_k)\|^2_{x_k} \geq 2\mu(f(x_k) - f(x_*))$$

and consequently, we have

$$f(x_{k+1}) - f(x_*) \leq \left(1 - \frac{\mu}{L}\right) \cdot (f(x_k) - f(x_*)) = \left(1 - \frac{1}{\kappa}\right) \cdot (f(x_k) - f(x_*))$$

as desired.

Now, since $x_k \in S_0$ for all $k \in \mathbb{N}_0$, since $x_* \in S_0$, and since S_0 is geodesically convex, there exists $v_k \in T_{x_*}M$ such that $x_k = \exp_{x_*}(v_k)$, and $[0, 1] \ni t \mapsto \exp_{x_*}(tv_k) \in S_0$. Consequently, by Eq. (2.4), we have

$$f(x_k) \geq f(x_*) + \langle \nabla f(x_*), v_k\rangle_{x_*} + \frac{\mu}{2}\|v_k\|^2_{x_*}$$

Since x_* is the minimizer of $f|_{S_0}$, by Corollary 2.9, $\nabla f(x_*) = 0$. Furthermore, $d(x_k, x_*) \leq \|v_k\|_{x_*}$, and thus

$$f(x_k) \geq f(x_*) + \frac{\mu}{2}\|v_k\|^2_{x_*} \geq f(x_*) + \frac{\mu}{2}d(x_k, x_*)^2$$

Thus

$$d(x_k, x_*) \leq \sqrt{\frac{2(f(x_k) - f(x_*))}{\mu}} \leq \sqrt{\frac{2(f(x_0) - f(x_*))}{\mu}} \cdot \sqrt{1 - \frac{1}{\kappa}}^k \tag{2.7}$$

Now, since M is complete, by the Hopf-Rinow theorem (Theorem A.2), there exists a length-minimizing geodesic γ such that $\gamma(0) = x_*, \gamma(1) = x_0, L(\gamma) = d(x_0, x_*)$. Furthermore, since f is geodesically convex on M,

$$f(\gamma(t)) \leq (1-t)f(\gamma(0)) + tf(\gamma(1)) = f(x_0) - (1-t)(f(x_0) - f(x_*)) \leq f(x_0)$$

for all $t \in [0, 1]$, and consequently, $\gamma([0, 1]) \subset S_0$. Applying Eq. (2.5) to γ (which can be viewed as an exponential map) yields (after recalling $\nabla f(x_*) = 0$) yields

$$f(x_0) \leq f(x_*) + \frac{L}{2} d(x_0, x_*)^2 \tag{2.8}$$

Combining Eq. (2.7) and Eq. (2.8) yields the desired result. ■

§3. Applications of Geodesic Convexity

3.1. Determining the Brascamp-Lieb Constant

Before we describe the applications of geodesic convexity, we take a brief detour into functional analysis and state the very important **Brascamp-Lieb inequality** ([BL76]):

Theorem 3.1 (Brascamp-Lieb Inequality)**.** Given linear maps $B = (B_j)_{j\in[m]}$, $B_j : \mathbb{R}^n \mapsto \mathbb{R}^{n_j}$, and non-negative real numbers $(p_j)_{j\in[m]}$, there exists a number $C \in [0, \infty]$ such that for *any* tuple of measurable functions $(f_j)_{j\in[m]}$, $f : \mathbb{R}^{n_j} \mapsto \mathbb{R}_{\geq 0}$, the following inequality holds:

$$\int_{x\in\mathbb{R}^n} \prod_{j=1}^{m} f_j(B_j x)^{p_j}\, dx \leq C \prod_{j=1}^{m} \left(\int_{x\in\mathbb{R}^{n_j}} f_j(x) dx \right)^{p_j}$$

The smallest C for which the above inequality holds is called the *Brascamp-Lieb constant* for the system (B, p), and is denoted as $\mathrm{BL}(B, p)$. A system (B, p) is called *feasible* if $\mathrm{BL}(B, p) < \infty$.
Bennett, Carbery, Christ, and Tao [BCCT08] showed that $\mathrm{BL}(B, p) < \infty$ if and only if the following criteria are satisfied:

1. $n = \sum_{j\in[m]} p_j n_j$.
2. $\dim(V) \leq \sum_{j\in[m]} p_j \dim(B_j V)$ for any subspace V of $\mathbb{R}^n$.

Henceforth, we will only be working with feasible Brascamp-Lieb systems.
Now, Lieb showed that equality occurs in the Brascamp-Lieb inequality when $f_j(x) = \exp(-x^\mathsf{T} A_j x)$ for some positive definite matrix A_j, for all $j \in [m]$. Plugging the above into the Brascamp-Lieb inequality yields:

$$\mathrm{BL}(B, p) \geq \left(\frac{\prod_{j\in[m]} \det(A_j)^{p_j}}{\det\left(\sum_{j\in[m]} p_j B_j^\mathsf{T} A_j B_j \right)} \right)^{1/2}$$

And thus

$$\mathrm{BL}(B, p) = \sup_{(X_1,\ldots,X_m)} \left(\frac{\prod_{j\in[m]} \det(X_j)^{p_j}}{\det\left(\sum_{j\in[m]} p_j B_j^\mathsf{T} X_j B_j \right)} \right)^{1/2}$$

where $X_j \in \mathbb{S}_{++}^{n_j}$ for all $j \in [m]$.
To simplify the expression a bit, we usually deal with the negative logarithm of it. Under that, we obtain:

$$-\ln \mathrm{BL}(B, p) = -\frac{1}{2} \sup_{(X_1,\ldots,X_m)} \left(\sum_{j\in[m]} p_j \ln\det(X_j) - \ln\det\left(\sum_{j\in[m]} p_j B_j^\mathsf{T} X_j B_j \right) \right)$$

Now, it can be shown that the function of $(X_1, \ldots, X_m)$ inside the supremum is not concave in the usual Euclidean sense: Indeed, suppose it was. Fix $X_2, \ldots, X_m$. Then we are effectively dealing with

$$p_1 \ln\det(X_1) - \ln\det\left(p_1 B_1^\mathsf{T} X_1 B_1 + C \right)$$

where C is some positive definite matrix. Note that $X_1 \mapsto \ln\det(X_1)$, and $X_1 \mapsto \ln\det\left(p_1 B_1^\mathsf{T} X_1 B_1 + C\right)$ are both concave functions and thus it is difficult to comment on the concavity of their difference. One might even suspect that the difference is not concave: That is indeed the case. There exist values of $X_2, \ldots, X_m$ for which the above

function is not concave. Refer to [VY18] for further details.
Thus, the usual tools of convex optimization fail for this problem. Here comes the true power of geodesics: We will prove that the above formulation is geodesically concave, and thus potentially amenable to methods of geodesic convex optimization.
Before coming to analyses of convexity, we prove another equivalent characterization of the Brascamp-Lieb constant:

$$-2\ln \mathrm{BL}(B,p) = \inf_{X\in\mathbb{S}^n_{++}} F_{B,p}(X)$$

where

$$F_{B,p}(X) := \sum_{j\in[m]} p_j \ln\det(B_j X B_j^\intercal) - \ln\det(X)$$

Theorem 3.2. $F_{B,p}$ is a geodesically convex function on the manifold $\mathbb{S}^n_{++}$ with the dot product induced by the Riemannian tensor being given by $g_X(U,V) := \mathrm{tr}(X^{-1}UX^{-1}V)$.

Proof. By Lemma 2.14, $\ln\det(X)$, and hence $-\ln\det(X)$ is geodesically linear, and hence geodesically convex. Thus it suffices to show that $\sum_{j\in[m]} p_j \ln\det(B_j X B_j^\intercal)$ is geodesically convex. Furthermore, since $p_j \geq 0$, it suffices to show that $\ln\det(B_j X B_j^\intercal)$ is geodesically convex. Equivalently, it suffices to show that $t \mapsto \ln\det(B_j M \exp(tN) M B_j^\intercal)$ is convex. But

$$\ln\det(B_j M \exp(tN) M B_j^\intercal) = \ln\det\exp(tN) + 2(\ln(\det(MB_j))$$

But $\det(\exp(tN)) = (\det\exp(N))^t$, and thus $\ln\det\exp(tN) = t\ln\det\exp(N)$, which is obviously convex. ∎

3.2. Operator Capacity For Square Operators

Another important problem in functional analysis is to find the *capacity* of a square operator, i.e. let T be a strictly positive linear operator. This problem has its origins in the so-called 'matrix-scaling problem' for bipartite graphs (see [Vis18]), but has since then grown to have applications in various fields of computer science and math, including 'non-commutative identity testing' (i.e. testing if a symbolic matrix of non-commuting variables over $\mathbb{Q}$ is invertible or not)! See [Gur04, GGOW16, GGOW17, AZGL+18] for further details.
Define:

$$\mathrm{cap}(T) := \inf_{X\in\mathbb{S}^n_{++}} \frac{\det(T(X))}{\det(X)}$$

It can be shown that the function $X \mapsto \ln\mathrm{cap}(X)$ is not convex. However, it is geodesically convex:

Lemma 3.3. $\mathbb{S}^n_{++} \ni X \mapsto \ln\mathrm{cap}(X)$ is a geodesically convex function.

Proof. Note that

$$\ln\mathrm{cap}(X) = \ln\det(T(X)) - \ln\det(X)$$

The first term is geodesically convex by Lemma 2.15, and $-\ln\det(X)$ is geodesically convex by Lemma 2.14. ∎

References

[AZGL+18] Zeyuan Allen-Zhu, Ankit Garg, Yuanzhi Li, Rafael Oliveira, and Avi Wigderson. Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 172–181, New York, NY, USA, 2018. Association for Computing Machinery. doi:10.1145/3188745.3188942.

[BCCT08] Jonathan Bennett, Anthony Carbery, Michael Christ, and Terence Tao. The brascamp-lieb inequalities: Finiteness, structure and extremals. *Geometric and Functional Analysis*, 17(5):1343–1415, 2008. doi:10.1007/s00039-007-0619-6.

[BL76] Herm Jan Brascamp and Elliott H Lieb. Best constants in young's inequality, its converse, and its generalization to more than three functions. *Advances in Mathematics*, 20(2):151–173, 1976. URL: https://www.sciencedirect.com/science/article/pii/0001870876901845, doi:10.1016/0001-8708(76)90184-5.

[Bou23] Nicolas Boumal. *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023. URL: https://www.nicolasboumal.net/book, doi:10.1017/9781009166164.

[GGOW16] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. A deterministic polynomial time algorithm for non-commutative rational identity testing. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 109–117, 2016. doi:10.1109/FOCS.2016.95.

[GGOW17] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. Algorithmic and optimization aspects of brascamp-lieb inequalities, via operator scaling. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 397–409, New York, NY, USA, 2017. Association for Computing Machinery. doi:10.1145/3055399.3055458.

[Gur04] Leonid Gurvits. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences*, 69(3):448–484, 2004. Special Issue on STOC 2003. URL: https://www.sciencedirect.com/science/article/pii/S0022000004000893, doi:10.1016/j.jcss.2004.06.003.

[Gur06] Leonid Gurvits. Hyperbolic polynomials approach to van der waerden/schrijver-valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '06, pages 417–426, New York, NY, USA, 2006. Association for Computing Machinery. doi:10.1145/1132516.1132578.

[SV17] Damian Straszak and Nisheeth K. Vishnoi. Real stable polynomials and matroids: optimization and counting. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 370–383, New York, NY, USA, 2017. Association for Computing Machinery. doi:10.1145/3055399.3055457.

[UDR77] CONSTANTIN UDRIŞTE. Continuity of convex functions on riemannian manifolds. *Bulletin mathématique de la Société des Sciences Mathématiques de la République Socialiste de Roumanie*, 21 (69)(1/2):215–218, 1977. URL: http://www.jstor.org/stable/43680306.

[Vis18] Nisheeth K. Vishnoi. Geodesic convex optimization: Differentiation on manifolds, geodesics, and convexity, 2018. arXiv:1806.06373.

[VY18] Nisheeth K. Vishnoi and Ozan Yildiz. On geodesically convex formulations for the brascamp-lieb constant. In *International Workshop and International Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2018. URL: https://api.semanticscholar.org/CorpusID:4794652.

§A. Appendix

A.1. Matrix Calculus

We include a few useful identities from matrix calculus for the reader's reference:

$$\frac{d\,\mathrm{tr}(X)}{dy} = \mathrm{tr}\left(\frac{dX}{dy}\right) \tag{A.1}$$

$$\frac{d(X^{-1})}{dy} = -X^{-1} \cdot \frac{dX}{dy} \cdot X^{-1} \tag{A.2}$$

$$\frac{de^{At}}{dt} = Ae^{At} = e^{At}A \tag{A.3}$$

A.2. The Exponential Map

Definition A.1 (Maximal Geodesics)**.** Let M be a Riemannian manifold. Then for every $(x, v) \in TM$, there exists a unique maximal geodesic $\gamma_{x,v} : I \mapsto M$ where I is an interval in $\mathbb{R}$ containing 0, such that $\gamma_{x,v}(0) = x, \gamma'_{x,v}(0) = v$. The maximality of I just means that there is no interval $I' \supsetneq I$ such that $\gamma_{x,v}$ can be extended to a geodesic (satisfying the given conditions) defined on I'.

Definition A.2 (Exponential Map)**.** Consider the following subset of the tangent bundle:

$$\mathcal{O} := \{(x, v) \in TM : \gamma_{x,v} \text{ is defined on an interval containing } [0, 1]\}$$

Also, consider its restriction at x:

$$\mathcal{O}_x := \{v \in T_xM : (x, v) \in \mathcal{O}\}$$

Then we define the exponential map as $\exp : \mathcal{O} \mapsto M$ as:

$$\exp(x, v) := \exp_x(v) := \gamma_{x,v}(1)$$

Remark. A few remarks are due:

1. A Riemannian manifold is called geodesically complete if $\mathcal{O} = TM$. In other words, the domain of $\gamma_{x,v}$ for *any* $(x, v) \in TM$ is $\mathbb{R}$.
2. Given $t \in \mathbb{R}$, if $tv \in \mathcal{O}_x$ for some v, then $\gamma_{tv}(1) = \gamma_v(t)$, i.e. $\exp_x(tv) = \gamma_v(t)$.
3. Fix any $x \in M$. If $v \in \mathcal{O}_x$, then $tv \in \mathcal{O}_x$ for all $t \in [0, 1]$. Consequently, $\mathcal{O}_x$ is *star-shaped* (a subset S of a $\mathbb{R}$-vector space is called star-shaped if $tS \subset S$ for all $t \in [0, 1]$).
4. $\mathcal{O}_x$ is open in T_xM, and hence $\mathcal{O}$ is open in TM. Note that $(x, 0) \in \mathcal{O}$ for all x. Consequently, $\mathcal{O}$ is a *neighborhood of the zero section of the tangent bundle.*
5. $\exp$ is smooth.
6. Fix $(x, v) \in \mathcal{O}$, and let $\gamma = \gamma_{x,v}|_{[0,1]}$. Then $L(\gamma) = \|v\|_x$.

A.3. Hopf-Rinow Theorem

A manifold M is called connected if it is connected as a topological space.

Lemma A.1. Connected manifolds are path-connected, i.e. if M is a connected manifold, then for any $x, y \in M$, there exists a continuous function $\tau : [0,1] \mapsto M$ with $\tau(0) = x, \tau(1) = y$.

Proof. Fix arbitrary $x \in M$, and denote as $U_x := \{y \in M : \text{ There is a path from } x \text{ to } y\}$. Note that $x \in U$, and hence $U \neq \varnothing$. We claim that U is open: Indeed, suppose $q \in U$, and let (V, ψ) be a coordinate chart such that $q \in V$. WLOG $\psi(V)$ is an open ball in an Euclidean space and hence is path-connected. Since V and $\psi(V)$ are homeomorphic, V is path connected. But that implies $V \subset U$: Indeed, if $v \in V$, then there is a path from x to v via q. Since V is open, we get that there is a neighborhood of q in U, showing that U is open. Now, note that if $\alpha \in U_\beta$, then $\beta \in U_\alpha$, and consequently, if $U_\alpha \neq U_\beta$, then $U_\alpha \cap U_\beta = \varnothing$. Finally, if $U_x \neq M$, then we could partition M into open sets as

$$M = \bigsqcup U_\alpha$$

which would contradict the fact that M was connected. ■

Remark. The above proof has been reproduced from here.

Thus, let M be a connected Riemannian manifold. Then, for any $x, y \in M$, we can define:

$$d(x,y) := \inf_{\substack{\tau:[0,1]\mapsto M \\ \tau \text{ is a path from } x \text{ to } y}} L(\tau)$$

where recall that

$$L(\tau) := \int_0^1 \sqrt{\langle \dot{\tau}(s), \dot{\tau}(s) \rangle_{\tau(s)}} ds$$

It is easy to verify that $d(\cdot, \cdot)$ turns M into a metric space.
The following is a very important theorem about connected Riemannian manifolds:

Theorem A.2 (Hopf-Rinow Theorem)**.** Let M be a connected Riemannian manifold. Then the following statements are equivalent:

1. The closed bounded sets of M are compact.
2. M is a complete metric space.
3. M is geodesically complete.

Furthermore, any one of the above implies the existence of a length-minimizing geodesic between any two given points on the manifold.

Remark. A few remarks are due:

1. The Hopf-Rinow theorem is not true for pseudo-Riemannian manifolds: See the Clifton-Pohl torus.
2. If a Riemannian manifold is not complete, there may not exist a geodesic between two given points in the first place: For example, consider the Riemannian manifold $\mathbb{R}^2 \setminus \{(0,0)\}$ embedded in $\mathbb{R}^2$, inheriting the usual smoothness and metric tensor. Then there is no geodesic between $(1,1)$ and $(-1,-1)$.